KOLOSO | Policy Note PN/2026/01

Evidence & Impact Series

Responsible AI in the Classroom:

How Koloso uses Artificial Intelligence, and why the design choices matter

James Blewett, Co-Founder and CEO, Koloso Ltd | March 2026 | koloso.app/evidence

Abstract

Artificial intelligence is becoming a standard feature label in educational technology products. But the label tells school leaders and teachers almost nothing about what the AI actually does, what data it processes, or what happens when it produces poor outputs. These are important questions, on which child safeguarding, data protection compliance, and pedagogical integrity depend.

This policy note sets out, in specific and verifiable terms, how Koloso uses AI in its Teaching Support System. It describes three design principles: 1) bounded scope; 2) teacher-gated outputs; and 3) zero-retention anonymised data architecture; that together define Koloso's approach to responsible AI deployment in African primary and secondary schools. It argues that these principles are not compliance minimums: they are the conditions under which AI can be genuinely useful in classrooms without adding new risks for the educators and institutions responsible for children's safety and learning.

1. The problem with the AI label in edtech

In early 2026, Kip Glazer — a US school administrator, author, and widely followed voice in educational technology — published an open letter to the edtech industry. She described attending a product demonstration in which she saw a button labelled 'AI-powered.' When she asked the vendor which large language model it was built on, she received the answer: 'I am not sure. I will have to get back to you.' She never reached her next questions: was the model being trained on student data, and had users consented to that?

Glazer was explicit that she was not asking from hostility to the technology. She wrote a book about the potential of generative AI in education. She described herself as someone who believes in the power of well-designed technology. Her point was simpler and sharper than scepticism:

> *"When you put GenAI on your product without being able to explain it clearly, you are not making my job easier. You are adding a new layer of risk I now have to manage."*
>
> Kip Glazer, LinkedIn, March 2026

Glazer proposed three questions that she believes every edtech company should be able to answer before labelling a product with AI:

- What does the AI actually do in your product, specifically?
- What data was it trained on, and did your users consent to that?

- What happens inside your system when the AI gets it wrong?

These are the right questions, ones that Koloso can answer precisely because the answers were built into the product architecture from the outset, not retrofitted in response to scrutiny. This policy note provides our answers, explains the design principles behind them, and sets out why we believe this approach should become a sector standard rather than a differentiator.

2. What Koloso's AI actually does: Four bounded tasks

Koloso uses Anthropic's Claude, accessed via API, to perform four specific, bounded tasks within its Teaching Support System. None of them involve the AI interacting directly with students.

Task	What the AI does
Curriculum extraction	When a school administrator uploads a national curriculum document, the AI reads the PDF and extracts its structure — year groups, domains, topics, and learning objectives — into the Koloso database. This replaces weeks of manual data entry.
Scheme of Work generation	AI allocates curriculum topics across the school calendar, respecting mandated time allocations. Input: curriculum content and term dates. Output: a week-by-week topic plan, for teacher review.
Lesson Plan generation	AI drafts lesson plans from the Scheme of Work, including suggested starter activities, main teaching approaches, differentiation strategies, and assessment prompts. Teachers review and edit before any plan is used.
Adaptive strategy generation	When assessment data reveals that a class, group, or individual student is struggling with specific learning objectives, AI generates a suggested teaching strategy: a diagnosis, recommended approach, supporting activities, and check questions. The teacher decides whether to adopt, adapt, or dismiss it.

The critical structural constraint: **the AI is never student-facing.** It generates drafts and suggestions for professional educators to review. Every AI output passes through a human gate:

- The teacher must actively adopt a plan or strategy before it enters the teaching workflow;
- A lesson plan the teacher has not reviewed does not reach students; and
- An adaptive strategy the teacher has not approved does not change what happens in the classroom.

This is not a feature of the user interface. It is a structural constraint in the system architecture. In Koloso, AI outputs are proposals, not actions.

3. Three design principles for responsible AI deployment

Koloso's approach to AI in education rests on three principles. Each is described below with the specific implementation choices that give it substance.

1 Bounded AI

Specific tasks with explicit scope limits — not general intelligence in the classroom

The most important decision in deploying AI responsibly is deciding what it will not do. Koloso's AI performs exactly the four tasks listed in Section 2. It does not generate feedback to students, assess student work, communicate with parents, make promotion or streaming decisions, or take any action in the system without teacher review.

This boundary matters because edtech AI risk is not uniform. The risk profile of AI that drafts a lesson plan for a teacher to review is fundamentally different from AI that assigns a student to an intervention group, generates a report for a parent, or flags a student's performance to school leadership. Koloso's AI does the former. It is structurally incapable of the latter.

This is not a limitation Koloso plans to remove as the technology matures. It is a principle Koloso intends to maintain and defend as a condition of responsible deployment in schools that serve children.

2 Teacher-gated AI

Every output is a suggestion — the teacher is always the decision-maker

In Koloso's Adapt module, when the system identifies a student, group, or class struggling with a specific learning objective, it generates an adaptive teaching strategy. That strategy arrives in the teacher's interface with status: Suggested. It does nothing until the teacher acts.

The teacher has three options:

- **Adopt:** The strategy enters the teaching workflow as planned.
- **Adopt with notes:** The teacher amends the AI's recommendation with their own professional judgement before implementing it.
- **Dismiss:** The teacher rejects the strategy, with an optional reason logged for system improvement.

Teachers can also request a regenerated strategy, asking the system to produce a different approach to the same diagnostic. A teacher who works with 60 students in a peri-urban school in Lusaka needs a different strategy from one who teaches 28 students in Lagos. The AI's first suggestion may not fit the context. The teacher's ability to dismiss and regenerate is the intended workflow not a workaround.

After a strategy is adopted and a follow-up assessment is completed, the system compares pre-scores with post-scores and rates the strategy's effectiveness: Effective, Partially Effective, or Ineffective. Ineffective strategies are surfaced to the teacher and logged, not hidden or averaged away. The teacher and the system both learn whether the AI's advice worked, and the feedback loop is closed.

3 Zero-retention, anonymised data architecture

The AI never sees a child's name or individual assessment record

This is the most technically specific of the three principles, and the most important for data protection and child safeguarding.

Koloso uses Anthropic's Claude via API under Anthropic's commercial terms, which explicitly state that customer data is not used for model training. Koloso is calling a pre-trained model, not fine-tuning one on school data. No student, teacher, or school data of any kind has contributed to the AI model's training.

In addition, Koloso sets the `anthropic-no-store` header on every API call — across all four AI tasks. This header instructs Anthropic not to retain or log any prompt or completion data, even for abuse monitoring. The practical result: no input Koloso sends to the AI, and no output the AI generates, is stored on Anthropic's infrastructure beyond the duration of the API call itself.

For adaptive strategy generation — the AI task most directly connected to individual student performance — the data architecture provides a further layer of protection. When the system constructs a prompt for the AI, it uses anonymised labels: 'Student 1,' 'Student 2,' and so on. Real student names are never included in the AI prompt. They are resolved only after the AI response is returned, matched locally in the school's own database. The AI model literally never processes a child's name or a named individual's assessment history.

What the AI does receive, by task:

- **Curriculum extraction:** Publicly available national curriculum documents, no school or student data.
- **Scheme of Work and Lesson Plan generation:** Curriculum content and school calendar structure, no student data.
- **Adaptive strategy generation:** Aggregated statistical data (average scores, performance zone distributions, and common error patterns) not individual assessment records.

4. What happens when the AI gets it wrong

Glazer's third question ("What happens inside your system when the AI gets it wrong?") is in some ways the most important, because AI will produce poor outputs. The question is not whether errors occur; it is whether the system is designed to contain them.

Koloso's architecture assumes AI imperfection from the outset. Every design decision described in this paper (bounded scope, teacher-gated outputs, anonymised data) is also an error-containment mechanism. If the AI produces a poor lesson plan, the teacher edits or discards it before it reaches students. If the AI produces a poor adaptive strategy, the teacher dismisses it and requests another.

The practical consequence of an AI error in Koloso's current architecture is this: a teacher reads a suggestion they disagree with, taps Dismiss, and moves on. The negative impact of an AI error is a teacher spending ten seconds reading a bad suggestion, which is then dismissed. It does not reach students, affect assessment records, or generate a report or an alert.

This is what we mean by 'teacher-gated' in practice.

The effectiveness measurement loop adds a further layer of accountability. If the AI's adaptive strategies for a particular learning objective consistently underperform across multiple teachers and schools, that pattern is visible to the Koloso product team. Systematic AI errors leave a traceable record. Bad patterns are surfaced, not hidden.

> **In short**
>
> The AI never contacts students, never assigns work, and never grades. It advises teachers, who remain the decision-makers. When it advises poorly, the teacher dismisses the advice. When it advises poorly and consistently, the system records and surfaces that pattern. This is what it means for AI to be bounded, auditable, teacher-gated, and zero-retention as a specific set of design choices with known and verifiable consequences.

5. Why this matters in African classroom contexts

Koloso operates in Zambia, South Africa, Nigeria, and Uganda. The schools it serves range from well-resourced urban private schools to under-resourced peri-urban government schools with large class sizes, intermittent connectivity, and teachers carrying significant administrative burdens alongside their teaching work.

The responsible AI principles described in this paper are designed for contexts where:

- Teachers may have 60 or more students in a single class, making personalised assessment analysis manually impossible at any useful frequency.
- School leaders are accountable to parents and government inspectors for teaching quality, but lack the data infrastructure to monitor it systematically.
- Data protection legislation is developing but enforcement is uneven, making voluntary compliance architecture more important, not less.

- Parents have limited visibility into their child's learning progress, and existing reporting mechanisms are infrequent and often unreliable.

In this context, AI that is genuinely useful is not a luxury and can make meaningful contributions to teaching quality and student outcomes - it can save a teacher hours of lesson planning, surface which students are falling behind before the end-of-year examination, and generate a specific teaching strategy for a group struggling with fractions.

But that usefulness only holds if the AI is deployed in a way that school owners, head teachers, teachers, and parents can trust. Trust in this context is not an abstract value: it is the condition under which a school subscribes, a teacher engages, and a parent allows their child's learning data to be processed. The architecture described in this paper is Koloso's answer to how to earn and sustain that trust.

6. A proposed framework for the sector

Koloso publishes this policy note not only to account for its own practices, but to propose that the three principles described here: 1) bounded scope; 2) teacher-gated outputs; and 3) zero-retention anonymised data architecture; should form a minimum standard for AI deployment in K–12 edtech products serving children in Zambia and beyond.

We recognise that not every AI use case in education fits this framework directly. Generative AI tutoring systems, automated essay scoring, and AI-driven student wellbeing monitoring all present different risk profiles and may require different design responses, and we do not claim that Koloso's framework is universal. However, we do claim that it is a useful starting point for a conversation the sector needs to have more explicitly.

Specifically, we propose that edtech companies deploying AI in schools serving children should be able to provide clear, verifiable answers to the following questions, and that school procurement processes should require those answers before contracts are signed:

- **What specific tasks does the AI perform?** Not 'AI-powered learning support' — what does it actually do, step by step?
- **Is any student data used to train or fine-tune the model?** Including anonymised or aggregated data. If so, what was the consent process?
- **Is student data retained by the AI provider beyond the API call?** What specific data retention settings are configured, and how are they verified?
- **Can a teacher override or dismiss every AI output before it affects students?** Is this a structural constraint in the system, or a user interface option that can be bypassed?
- **How are AI errors detected, recorded, and surfaced?** What is the mechanism by which a systematically poor AI output is identified and corrected?

These are the questions Kip Glazer was raising, and they are the questions that school leaders, parents, and regulators will increasingly require edtech companies to answer. Our belief is that companies that can answer them clearly and specifically will build durable trust, and companies that cannot will find that the AI label, deployed without substance, becomes a liability rather than a differentiator.

7. Summary

What Koloso's AI does	Four bounded tasks: curriculum extraction, Scheme of Work generation, Lesson Plan generation, and adaptive strategy generation. The AI is never student-facing.
What it does not do	Contact students, assign work, grade work, generate parent communications, or take any action in the system without teacher review.
Training data	Koloso calls a pre-trained model (Anthropic Claude) via API. No student, teacher, or school data has been used for model training. Anthropic's commercial API terms explicitly prohibit customer data from being used for training.
Data retention	The anthropic-no-store header is set on every API call. No prompt or completion data is retained by Anthropic beyond the duration of the API call.
Student PII in prompts	Never included. Adaptive strategy prompts use anonymised labels (Student 1, Student 2). Real names are resolved locally, after the AI response is returned. The AI model never processes a child's name.
Error containment	Every AI output arrives with status 'Suggested.' Teachers adopt, adopt with notes, dismiss, or regenerate. Ineffective strategies are logged and surfaced, not hidden. The maximum impact of an AI error is a teacher reading a poor suggestion and tapping 'Dismiss.'

Disclosure and limitations

This policy note is authored by the CEO of Koloso Ltd and describes Koloso's own product architecture. Readers should apply appropriate scrutiny to the claims made. The technical claims regarding the Anthropic API — including the commercial data retention terms and the anthropic-no-store header — are verifiable against Anthropic's published documentation and API reference. Koloso's commitment to transparency on AI use is ongoing: this policy note will be updated to reflect any material changes to Koloso's AI architecture, provider, or data practices. Questions or requests for technical clarification should be directed to james@koloso.app.

Koloso Ltd | PN/2026/01 | Responsible AI in Edtech | March 2026 | koloso.app/evidence | james@koloso.app